



## Causality, Modality and Explanation

White, Graham

For additional information about this publication click this link.

<http://qmro.qmul.ac.uk/jspui/handle/123456789/5063>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact [scholarlycommunications@qmul.ac.uk](mailto:scholarlycommunications@qmul.ac.uk)

ISSN 1470-5559

# Causality, Modality, and Explanation

Graham White



RR-06-03

April 2006





# Causality, Modality, and Explanation

Graham White  
Department of Computer Science  
Queen Mary, University of London  
London E2 9QD  
graham@dcs.qmul.ac.uk  
<http://www.dcs.qmul.ac.uk/~graham>

April 29, 2006

## Abstract

We analyse the system of nonmonotonic logic invented by McCain and Turner, which is usually referred to as “causal reasoning”. We argue the following: that the McCain-Turner system can be perspicuously reformulated as a modal logic; that, so reformulated, it is best regarded as a logic of *explanation* rather than of causality; and, finally, that this logic of explanation has illuminating connections with a logic of argument due to Parsons and Jennings.

In the course of our argument, we give a sequent calculus formulation of McCain and Turner’s logic, and show that it satisfies cut elimination.

## Contents

<b>1</b>	<b>Motivation</b>	<b>1</b>
1.1	The McCain-Turner System . . . . .	2
1.2	The Road Ahead . . . . .	2
<b>2</b>	<b>The Modal Logic</b>	<b>3</b>
2.1	Cut Elimination . . . . .	6
2.2	The Structure of the Canonical Model . . . . .	8
<b>3</b>	<b>Explanation and Argument</b>	<b>12</b>

## 1 Motivation

This paper will start from a theme which is a little removed from the mainstream of formal logic, namely from the discussion, in the Artificial Intelligence community, of the so-called “frame problem”: this is the problem of reasoning, using logic, about action and change. Although this discussion has generated

a considerable quantity of theory, we will discuss here a single approach: the system invented by McCain and Turner, and which is usually referred to as “causal reasoning”.

### 1.1 The McCain-Turner System

The [7] treatment of the frame problem goes, roughly, as follows. They start with a binary sentential operator: let us write it  $\triangleright$ . Applications of this operator – of the form  $\phi \triangleright \psi$  – are called “causal rules”, and they can, in McCain and Turner’s treatment, be regarded as purely metatheoretic assertions. A collection of causal rules will be called a *causal theory*. Given a language  $\mathcal{L}$  and a causal theory  $\Xi$ , we define an operator  $(\cdot)^\Xi$ , from models to sets of sentences, as follows:

$$M^\Xi = \{\psi \mid \phi \triangleright \psi \in D, M \models \phi\}. \quad (1)$$

We now say that a model  $M$  of  $\mathcal{L}$  is *causally explained* iff it is the only model of  $M^\Xi$ . And we say that a proposition  $P$  is a consequence of  $\Xi$  iff it is true in all of the  $\Xi$ -causally explained models of  $\mathcal{L}$ . Given suitably chosen causal rules, this procedure appears to work: that is, it yields correct solutions of the frame problem, and it is fairly efficient computationally.

### 1.2 The Road Ahead

As it stands, though, this definition is not entirely perspicuous, either mathematically or conceptually. The definition of causally explained models is mysterious: we would, thus, like a more illuminating treatment of the mathematics. Furthermore, it is difficult to explain the conceptual role of  $\triangleright$ . For example, one of McCain and Turner’s standard causal rules is of the form

$$\phi(t) \wedge \phi(t+1) \triangleright \phi(t+1), \quad (2)$$

where  $\phi(t)$  means that  $\phi$  holds at time  $t$ . But reading  $\triangleright$  as ‘causes’ here is simply implausible: propositions – or the states of affairs which they denote – are not usually thought of as causing themselves.

So, we will do two things in this paper. Firstly, we will give an alternative definition of (1): it can easily be reformulated as a relation between models, and relations on a set of models give modal operators. So we can, instead, define a suitable modal operator,  $\Box$ ; given this operator, the deductive closure of (1) turns out to be

$$\{P \mid M \models \Box P\}, \quad (3)$$

and the relation

$$\phi \vdash \Box \psi.$$

is equivalent to  $\phi \triangleright \psi$  (equivalent in the sense that the two relations give the same set of explanatorily closed models).

But, as well as merely technical reformulations, we also want to say what these constructions *mean*. We argue, then, that a better reading of  $\phi \triangleright \psi$  is that

‘ $\phi$  explains  $\psi$ ’ – and, correspondingly,  $\Box\psi$  can be regarded as a disjunction of all of the possible explanations of  $\psi$ . In fact, this explanatory reading has been anticipated: Lifschitz paraphrases  $\phi \triangleright \psi$  as “[ $\psi$ ] has a cause if [ $\phi$ ] is true, or . . . [ $\phi$ ] provides a ‘causal explanation’ for [ $\psi$ ]” [4, p. 451]. And, given this explanatory reading, rules such as (2) seem far less contentious: (2) simply says that a good explanation for  $\phi$  being true at  $t + 1$  is that it was true at  $t$  and has the same truth values at  $t$  and  $t + 1$ . Furthermore this reading naturally allows certain facts to “explain themselves”, something which is quite usual in McCain-Turner style formalisms. Self-explanation is a good deal less problematic than self-causation, since explanations, after all, have to come to an end (or, of course, loop) at some point – see, for example [17, §217].

And we can go further than this. McCain and Turner’s theory seems, from this perspective, to be a very natural formulation of explanation *in general*: although its original application may have been to a causal context, there is nothing about it which forces these explanations to be *causal*. Once we broaden our horizons to general explanation, we can bring this modal system into contact with other work: for example, [8] (see also [9]) have a natural deduction system for formalising arguments, which system is a fragment of ours. And, where we have explanations, questions must also be in the neighbourhood: and thus we could also express a good deal of the formalism of [1] (see [3]) in terms of ours.

## 2 The Modal Logic

Our system will be given by a sequent calculus, as in [16]; it is given by the rules in Table 1, whose formulation depends on an underlying set of causal rules. A causal rule, here, is an ordered pair of propositions, written  $\phi \triangleright \psi$ :  $\phi$  and  $\psi$  are allowed to have free variables, but, in a given causal rule,  $\phi$  can have no more free variables than  $\psi$ .

This modal theory, as we will show, captures the meaning of the McCain-Turner causal implications, in the sense that a causal rule  $\phi \triangleright \psi$  can be expressed as  $\phi \vdash \Box\psi$ . In this it is somewhat similar to the system of [15], but, unlike Turner, we have a proof-theory with good properties.

We should remark that the rule for  $\Box L$  is, *in general*, infinitary (and even when finite, its set of premises can well be undecidable). The qualification “in general” is important here: typical applications will use sequents for which this rule always has a finite, decidable (and, indeed, very tractable) set of premises. Indeed, many applications will only use the right rule for  $\Box$ , which is much more tractable. Furthermore, even when the rule is actually infinitary, the system is still very useful for metatheoretical purposes: after all, the proof-theoretic use of infinitary systems has a very long and respectable history (see, for example, [2, p. 164]).

**Remark 1.** As it stands, the modal language extends the original language  $\mathcal{L}$  by the modal operator  $\Box$ : call the new language  $\mathcal{L}_\Box$ . However, cut elimination will show that  $\mathcal{L}_\Box$  is a definitional extension of  $\mathcal{L}$ : indeed, if we were to assume

Table 1: Sequent Calculus Rules

$\frac{}{A \vdash A} \text{Ax}$	$\frac{}{\perp \vdash} L \perp$
$\frac{\Gamma \vdash \Delta}{\Gamma, A \vdash \Delta} LW$	$\frac{\Gamma \vdash \Delta}{\Gamma \vdash A, \Delta} RW$
$\frac{\Gamma, A, A \vdash \Delta}{\Gamma, A \vdash \Delta} LC$	$\frac{\Gamma \vdash A, A, \Delta}{\Gamma \vdash A, \Delta} RC$
$\frac{\Gamma, A, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta} \wedge L$	$\frac{\Gamma \vdash A, \Delta \quad \Gamma \vdash B, \Delta}{\Gamma \vdash A \wedge B, \Delta} \wedge R$
$\frac{\Gamma, A \vdash \Delta \quad \Gamma, B \vdash \Delta}{\Gamma, A \vee B \vdash \Delta} \vee L$	$\frac{\Gamma \vdash A, B, \Delta}{\Gamma \vdash A \vee B, \Delta} \vee R$
$\frac{\Gamma \vdash A, \Delta \quad \Gamma, B \vdash \Delta}{\Gamma, A \rightarrow B \vdash \Delta} \rightarrow L$	$\frac{\Gamma, A \vdash B, \Delta}{\Gamma \vdash A \rightarrow B, \Delta} \rightarrow R$
$\frac{\Gamma, A[x/t] \vdash \Delta}{\Gamma, \forall x A \vdash \Delta} \forall L$	$\frac{\Gamma \vdash A[x/y], \Delta}{\Gamma \vdash \forall x A, \Delta} \forall R$
$\frac{\Gamma, A[x/y] \vdash \Delta}{\Gamma, \exists x A \vdash \Delta} \exists L$	$\frac{\Gamma \vdash A[x/t], \Delta}{\Gamma \vdash \exists x A, \Delta} \exists R$
$\frac{\Gamma \vdash \phi_1 \wedge \dots \wedge \phi_n, \Delta \quad \psi_1, \dots, \psi_n \vdash X}{\Gamma \vdash \Box X, \Delta} \Box R$	
$\frac{\left\{ \Gamma, \phi_{i_1}, \dots, \phi_{i_{k_i}} \vdash \Delta, \quad \psi_{i_1}, \dots, \psi_{i_{k_i}} \vdash X \right\}_{i \in I}}{\Gamma, \Box X \vdash \Delta} \Box L$	
$\frac{\Gamma \vdash X^m, \Delta \quad \Gamma', X^n \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta, \Delta'} \text{multicut}$	

**conditions on the rules:**

- $\forall R$   $y$  not free in  $\Gamma$  or  $\Delta$ , and either  $y = x$  or  $y$  not free in  $A$
- $\exists L$   $y$  not free in  $\Gamma$  or  $\Delta$ , and either  $y = x$  or  $y$  not free in  $A$
- $\Box R$  where, for all  $i$ ,  $\phi_i \triangleright \psi_i$  is an instantiation of a causal rule.
- $\Box L$  for each appropriate  $i$  and  $j$ , we have an instantiation of a causal rule  $\phi_{i_j} \triangleright \psi_{i_j}$ , and where the  $\{\phi_{i_j}\}$  and  $\{\psi_{i_j}\}$ , for  $i \in I$  and, for each  $i$ ,  $j = 1, \dots, k_i$ , are the only such sets of  $\phi$ s and  $\psi$ s that there are

multicut  $X^n$  stands for  $n$  occurrences of  $X$ ;  $m, n > 0$

that  $\mathfrak{L}$  has arbitrary disjunctions, we could write

$$\Box P = \bigvee_{\psi_1, \dots, \psi_k \vdash P} \phi_i \wedge \dots \wedge \phi_k$$

So,  $\mathfrak{L}_\Box$  is rather closely tied to  $\mathfrak{L}$ : the following proposition gives an illuminating characterisation of  $\mathfrak{L}_\Box$ .

**Proposition 1.** *In the presence of arbitrary disjunctions,  $\Box$  is the minimal **K** modality which satisfies*

$$\phi \vdash \Box \psi \quad \text{for all } \phi \triangleright \psi :$$

to be precise, suppose that  $\mathfrak{L}_{\Box'}$  is an extension of  $\mathfrak{L}$  by a modal operator  $\Box'$ , such that

- $\mathfrak{L}_{\Box'}$  has arbitrary disjunctions,
- there is an entailment relation  $\vdash'$  on  $\mathfrak{L}_{\Box'}$ ,
- relative to this entailment relation,  $\Box'$  is a **K** modality, and
- for any  $\phi \triangleright \psi$ , we have  $\phi \vdash' \Box' \psi$ .

Then there is an interpretation  $\alpha : \mathfrak{L}_\Box \rightarrow \mathfrak{L}_{\Box'}$ , compatible with the entailment relations, such that, for all  $P \in \mathfrak{L}_\Box$ ,

$$\alpha(\Box P) \vdash' \Box' \alpha(P).$$

*Proof.* Define  $\alpha$  as follows. It is the identity on non-modal propositions, and its effect on modalisations is given by

$$\alpha(\Box P) = \bigvee_{\psi_1, \dots, \psi_k \vdash P} \phi_1 \wedge \dots \wedge \phi_k$$

(the right hand side is clearly an element of  $\mathfrak{L}$  and so is in  $\mathfrak{L}_{\Box'}$ ); extend  $\alpha$  truth-functionally to the whole of  $\mathfrak{L}_\Box$ . We now show that  $\alpha$  is compatible with the entailment relations: we do this by supposing that  $\Gamma \vdash \Delta$ , and showing, by induction on the proof tree, that  $\alpha(\Gamma) \vdash' \alpha(\Delta)$ . This is clear for axioms, and for the non-modal rules: we now consider the modal rules. For  $\Box R$ , this comes down to showing that, if  $\alpha(\Gamma) \vdash \phi_1 \wedge \dots \wedge \phi_k$ ,  $\alpha(\Delta)$ , and if  $\psi_1, \dots, \psi_k \vdash P$ , then  $\alpha(\Gamma) \vdash \alpha(\Box P)$ ,  $\alpha(\Delta)$ ; but this follows immediately from the definition of  $\alpha(\Box P)$ . The proof of the case for  $\Box L$  is similarly trivial.

Now we show that, for any  $P \in \mathfrak{L}_\Box$ ,

$$\alpha(\Box P) \vdash' \Box' \alpha(P)$$

We apply the definition of  $\alpha$  and the left rule for  $\bigvee$ : we then have to prove that, if  $\psi_1, \dots, \psi_k \vdash P$ , that  $\phi_1, \dots, \phi_k \vdash \Box' \alpha(P)$ . But this follows from the properties of **K** modalities and the assumptions made on  $\Box'$ .  $\square$

After our cut elimination result, we can prove a version of Proposition 1 which does not rely on the existence of arbitrary disjunctions.



## 2.1 Cut Elimination

**Theorem 1.** *Given a McCain-Turner causal theory, the corresponding modal system satisfies cut elimination.*

*Proof.* (Cf. [16]) This is analogous to the theorem of [13]: however, because the system is, in general, infinitary, proof trees may contain branches of unbounded length, and we have to be careful about the induction we use.

The interesting case is when we have a cut of  $\Box R$  against  $\Box L$ , and where the cutformulae are principal on both sides. We have:

$$\frac{\frac{\Gamma \vdash \phi_1 \wedge \dots \wedge \phi_k, (\Box X)^m \Delta}{\Gamma \vdash (\Box X)^{m+1}, \Delta} \Box R \quad \frac{\left\{ \frac{\frac{\Pi_i}{\vdots} \Gamma', \phi_{i_1}, \dots, \phi_{i_{k_i}}, (\Box X)^n \vdash \Delta'}{(\Box X)^{n+1} \vdash \Delta'} \Box L \right\}_{i \in I}}{\Gamma', (\Box X)^{n+1} \vdash \Delta'} \text{multicut}}{\Gamma, \Gamma' \vdash \Delta, \Delta'}$$

We also suppose that  $m, n > 0$ ; other cases are similar but simpler.

We first move the cut up on the left:

$$\frac{\frac{\Gamma \vdash \phi_1 \wedge \dots \wedge \phi_k, (\Box X)^m \Delta}{\Gamma, \Gamma' \vdash \phi_1 \wedge \dots \wedge \phi_k, \Delta, \Delta'} \text{multicut} \quad \frac{\left\{ \frac{\frac{\Pi_i}{\vdots} \Gamma' \phi_{i_1}, \dots, \phi_{i_{k_i}}, (\Box X)^n \vdash \Delta'}{(\Box X)^{n+1} \vdash \Delta'} \right\}}{\Gamma', (\Box X)^{n+1} \vdash \Delta'}}$$

Now the cut has been moved up on the left, and the cutformula is the same: so we may inductively apply cut elimination to this proof, and obtain a cutfree proof (call it  $\mathcal{P}_0$ ) of its result.

We next note that the set  $\phi_1, \dots, \phi_k$  on the left must correspond to one of the sets  $\phi_{i_1}, \dots, \phi_{i_{k_i}}$  on the right, so that we can move the cut up on the right:

$$\frac{\frac{\frac{\Pi}{\vdots} \Gamma \vdash \phi_1 \wedge \dots \wedge \phi_n, (\Box X)^m, \Delta}{\Gamma \vdash (\Box X)^{m+1}, \Delta} \quad \frac{\frac{\frac{\Pi_{i_0}}{\vdots} \Gamma', \phi_1, \dots, \phi_k, (\Box X)^n \vdash \Delta}{\Gamma, \Gamma', \phi_1, \dots, \phi_k \vdash \Delta, \Delta'} \text{multicut}}{\Gamma, \Gamma', \phi_1, \dots, \phi_k \vdash \Delta, \Delta'}$$

for a suitable  $i_0$ ; again we have moved the cut up so we can, inductively, obtain a cut free proof (call this  $\mathcal{P}_1$ ) of its result.

Finally we put these proofs together: we get

$$\begin{array}{c}
\begin{array}{c} \mathcal{P}_0 \\ \vdots \\ \Gamma, \Gamma' \vdash \phi_1 \wedge \dots \wedge \phi_k, \Delta, \Delta' \end{array} \quad \begin{array}{c} \mathcal{P}_1 \\ \vdots \\ \Gamma, \Gamma', \phi_1, \dots, \phi_k \vdash \Delta, \Delta' \\ \vdots \\ \wedge R \\ \Gamma, \Gamma', \phi_1 \wedge \dots \wedge \phi_k \vdash \Delta, \Delta' \end{array} \\
\hline
\frac{\Gamma, \Gamma' \vdash \phi_1 \wedge \dots \wedge \phi_k, \Delta, \Delta' \quad \Gamma, \Gamma', \phi_1 \wedge \dots \wedge \phi_k \vdash \Delta, \Delta'}{(\Gamma)^2, (\Gamma')^2 \vdash (\Delta)^2, (\Delta')^2} \text{multicut} \\
\hline
\Gamma, \Gamma' \vdash \Delta, \Delta' \quad LC, RC
\end{array}$$

This cut has not been moved upwards, but the cutformula has been changed from  $\Box X$  to  $\phi_1 \wedge \dots \wedge \phi_k$ ; although the logical complexity may have increased, the modal depth will (because of our restriction on the form of the causal rules) have decreased.

This is the essential idea of the proof: however, because we have an infinitary system, we have to use the methods of ordinal analysis (see, for example, Pohlers [11]) in order to carry out the induction.  $\square$

So, trivially:

**Corollary 1.** *The modal theory is a conservative extension of the non-modal theory.*

We also have our promised generalisation of Proposition 1. As in [10], we define the **K** rule to be

$$\frac{\Gamma \vdash A}{\Gamma', \Box \Gamma \vdash \Box A, \Delta'} \mathbf{K}$$

**Proposition 2.** *Suppose that  $\mathcal{L}_{\Box'}$  is an extension of  $\mathcal{L}$  by a modal operator  $\Box'$ , and suppose that we have an entailment relation  $\vdash'$  on  $\mathcal{L}_{\Box'}$  given by a sequent calculus with the **K** rule, together with a set of axioms closed under substitution. Suppose also that we have, for any  $\phi \triangleright \psi$ ,*

$$\phi \vdash' \Box' \psi$$

*Then there is a conservative extension  $\langle \tilde{\mathcal{L}}, \tilde{\vdash} \rangle$  of  $\langle \mathcal{L}_{\Box'}, \vdash' \rangle$ , and an interpretation  $\alpha$  of  $\mathcal{L}_{\Box}$  in  $\tilde{\mathcal{L}}$ , such that, for any  $P \in \mathcal{L}_{\Box}$ ,*

$$\alpha(\Box P) \tilde{\vdash} \Box' (\alpha P)$$

*Proof.* Extend  $\langle \mathcal{L}_{\Box'}, \vdash' \rangle$  to  $\langle \tilde{\mathcal{L}}, \tilde{\vdash} \rangle$  by adding another modality,  $\tilde{\Box}$ , together with rules for  $\tilde{\Box}$  of the same form as in Table 1. Then we can prove that proofs in  $\langle \tilde{\mathcal{L}}, \tilde{\vdash} \rangle$  can be transformed so as to have cuts only against axioms, as in [14, p. 97]. But this proves conservativity of  $\tilde{\mathcal{L}}$  over  $\mathcal{L}'$ , since the axioms are all in  $\mathcal{L}'$ . We now define an interpretation of  $\mathcal{L}_{\Box}$  in  $\tilde{\mathcal{L}}$ : it will be the identity on  $\mathcal{L}$ , and will interpret  $\Box$  by  $\tilde{\Box}$ . The rest of the proof is exactly the same as for Proposition 1.  $\square$

## 2.2 The Structure of the Canonical Model

We can now apply our cut elimination result to elucidate the McCain-Turner definitions; we recall the definition, (1), of the operation  $M^\Xi$ , from models to sets of sentences. The key result here is

**Lemma 1.** *If we have*

$$A \vdash \Box P$$

*in the sequent calculus, with  $A$  and  $p$  non modal, then there is a finite set*

$$I = \{\phi_{i_1} \wedge \dots \wedge \phi_{i_{k_i}}\}_{i \in I}$$

*of conjunctions of rule bodies such that, for each  $i \in I$ , the corresponding rule heads entail  $p$ , i.e.*

$$\psi_{i_1}, \dots, \psi_{i_{k_i}} \vdash P,$$

*and such that*

$$A \vdash I.$$

*Proof.* The proof is by an induction on the structure of cut-free proofs: because we may apply weakening or contraction to the occurrence of  $\Box P$  on the right, the inductive hypothesis will be:

If we have

$$\Gamma \vdash \Gamma', (\Box P)^n$$

in the sequent calculus, with  $n \geq 0$ , then there is a set

$$I = \{\phi_{i_1} \wedge \dots \wedge \phi_{i_{k_i}}\}_{i \in I}$$

of conjunctions of rule bodies such that, for each  $i \in I$ , the corresponding rule heads entail  $p$ , i.e.

$$\psi_{i_1}, \dots, \psi_{i_{k_i}} \vdash P,$$

and such that

$$\Gamma \vdash \Gamma', I.$$

The single-premise rules are trivial: the other cases are as follows. We will tacitly assume, in this proof, that, whenever we have a conjunction  $\phi_1, \dots, \phi_k$ , that the  $\phi_i$  are all instantiations of rule bodies, and that their heads satisfy  $\psi_1, \dots, \psi_k \vdash p$ . The letters  $I, J, \dots$  will be understood to refer to sets of such conjunctions of rule bodies.

**$\vee L$**  We start with

$$\frac{\Gamma_1, A \vdash \Gamma_2, (\Box P)^m \quad \Gamma_3, B \vdash \Gamma_4, (\Box P)^n}{\Gamma_1, \Gamma_3, A \vee B \vdash \Gamma_2, \Gamma_4, (\Box P)^{m+n}} \vee L$$

We prove the inductive hypothesis for the conclusion as follows:

$$\frac{\begin{array}{c} \Pi \\ \vdots \\ \Gamma_1, A \vdash \Gamma_2, I \end{array} \quad \begin{array}{c} \Pi' \\ \vdots \\ \Gamma_3, B \vdash \Gamma_4, J \end{array}}{\Gamma_1, \Gamma_3, A \vee B \vdash \Gamma_2, \Gamma_4, I, J}$$

where  $\Pi$  and  $\Pi'$  are the proofs given by the inductive hypothesis.

$\wedge R$  and  $\rightarrow L$  are similar.

$\Box R$  We have

$$\frac{\Gamma \vdash \Gamma', \phi_1 \wedge \dots \wedge \phi_k, (\Box P)^n}{\Gamma \vdash \Gamma', (\Box P)^{n+1}} \Box R$$

The next stage of the induction is as follows:

$$\frac{\begin{array}{c} \vdots \\ \Gamma \vdash \Gamma', \phi_1 \wedge \dots \wedge \phi_k, I \end{array}}{\Gamma \vdash \Gamma', I \cup \{\phi_1 \wedge \dots \wedge \phi_k\}} \text{structural rules}$$

**Axiom** We have

$$\frac{}{\Gamma, X \vdash X, \Gamma', (\Box P)^n} \text{Axiom}$$

and this immediately gives us the base case of the induction, with  $I = \emptyset$ .

$\perp L$  and  $\top R$  are similar.

□

**Lemma 2.** *Let  $\Xi$  be a causal theory, and let  $M$  be a model of the non-modal language: let  $\mathcal{M}$  be the set of (non-modal) sentences true in  $M$ . Then*

$$\overline{M^\Xi} = \{Q \mid Q \text{ nonmodal, and } \mathcal{M} \vdash \Box Q\} \quad (4)$$

where  $\overline{(\cdot)}$  is closure under the non-modal inference rules, and  $\vdash$  on the right is modal entailment.

*Proof.* We prove the right-to-left containment of (4) first. Suppose that we have a proof of  $\mathcal{M} \vdash \Box Q$ ; we can assume we have a cut-free proof of

$$P_1, \dots, P_k \vdash \Box Q$$

with the  $P_i$  true in  $M$ . (Note that, since the *right* rule for  $\Box$  is finitary, we can assume that the antecedent of this sequent is finite.) Then, by the previous lemma, there is a finite set  $I$  of conjunctions of rule bodies such that

$$P_1, \dots, P_k \vdash I$$

and such that all of the corresponding conjunctions of rule heads entail  $Q$ .

Because all of the  $P_i$  are true in  $M$ , at least one of the elements of  $I$  must be true in  $M$ : consequently, we have  $\phi_1, \dots, \phi_k$ , all true in  $M$ , such that

$$\psi_1, \dots, \psi_k \vdash Q$$

Hence  $Q \in \overline{M^\Xi}$ , and this proves the right-to-left inclusion.

The left-to-right direction of (4) is proved as follows. The right hand side is clearly deductively closed, so we only need to show that

$$M^\Xi \subseteq \{Q \mid Q \text{ nonmodal, and } \mathcal{I} \vdash \Box Q\}$$

So suppose that we have an element  $\psi$  of  $M^\Xi$ : we have instantiations of causal rules with head  $\psi$  and body  $\phi$ , and  $\phi$  is true in  $I$  (that is,  $\phi \in \mathcal{M}$ ). So we have a proof:

$$\frac{\frac{}{\phi \vdash \phi} \text{Ax}}{\phi \vdash \Box \psi} \Box R$$

So, consequently,  $\psi$  is a member of the right hand side.  $\square$

**Proposition 3.**  $\mathcal{L}_\Box$  is a definitional extension of  $\mathcal{L}$ : more precisely, for any model  $M$  of  $\mathcal{L}$ , if  $\mathcal{M}$  is the set of propositions true in  $M$ , and for any proposition  $Q$  of  $\mathcal{L}_\Box$ , we either have

$$\mathcal{M} \vdash Q$$

or

$$\mathcal{M} \vdash \neg Q$$

*Proof.* We prove this by induction on the modal complexity of  $Q$ : it is clearly true for  $Q$  unmodalised, since  $M$  is a model of  $\mathcal{L}$ , and since  $\vdash$  coincides with  $\vdash_\Box$  on  $\mathcal{L}$ . For the inductive step, we have to show that, for any  $\Box R$ , we either have  $\mathcal{M} \vdash \Box R$  or  $\mathcal{M} \vdash \neg \Box R$ . So,

$$\begin{array}{lll} & \mathcal{M} \not\vdash \neg \Box R & \\ \text{iff} & \mathcal{M}, \Box R \not\vdash \perp & \\ \text{iff} & \mathcal{M}, \phi_1 \wedge \dots \phi_k \not\vdash \perp & \text{for some } \phi_1, \dots, \phi_k \\ & & \text{with } \psi_1, \dots, \psi_k \vdash R \\ \text{iff} & \mathcal{M} \vdash \phi_1 \wedge \dots \phi_k & \text{since } M \text{ is a model} \\ \text{iff} & \mathcal{M} \vdash \Box R. & \end{array}$$

$\square$

**Theorem 2.** The canonical model of  $\mathcal{L}_\Box$  has the following structure: the worlds are all the models of  $\mathcal{L}$ , and the accessibility relation  $\rho$  is given by

$$I \rho J \text{ iff } J \text{ is a model of } I^\Xi$$

*Proof.* By definition, the worlds of the canonical model are simply the maximal consistent subsets of  $\mathfrak{L}_\square$ : because  $\mathfrak{L}_\square$  is a definitional extension of  $\mathfrak{L}$ , these are the same as maximal consistent subsets (i.e. models) of  $\mathfrak{L}$ . Lemma (4) gives the accessibility relation.  $\square$

**Lemma 3.** *Let  $I$  be a model of the non-modal language: then  $I$  is the only model of  $I^\Xi$  iff  $\mathcal{I} \vdash (\Box P \rightarrow P) \wedge (P \rightarrow \Box P)$ , for all non-modal  $P$ .*

*Proof.*  $I$  is the only model of  $I^\Xi$  iff

$$\begin{aligned} \mathcal{I} &= \overline{I^\Xi} \\ &= \{Q \mid Q \text{ nonmodal, and } \mathcal{I} \vdash \Box Q\} \end{aligned}$$

So  $\mathcal{I} \subseteq \overline{I^\Xi}$  iff

$$\mathcal{I} \vdash \Box Q \text{ if } I \models Q,$$

which is equivalent to

$$\mathcal{I} \vdash \Box Q \text{ if } \mathcal{I} \vdash Q,$$

and  $\overline{I^\Xi} \subseteq \mathcal{I}$  iff

$$I \models Q \text{ if } \mathcal{I} \vdash \Box Q$$

which is similarly equivalent to

$$\mathcal{I} \vdash Q \text{ if } \mathcal{I} \vdash \Box Q$$

Now, for any non-modal  $Q$ , we have  $\mathcal{I} \vdash Q$  or  $\mathcal{I} \vdash \neg Q$ ; so, finally, we have that  $I$  is the only model of  $I^\Xi$  iff

$$\mathcal{I} \vdash \Box Q \rightarrow Q \wedge Q \rightarrow \Box Q$$

for any non-modal  $Q$ .  $\square$

This allows us to prove

**Theorem 3.** *Given a McCain-Turner causal theory  $\Xi$ , a non-modal proposition  $P$  is causally explained according to  $\Xi$  if, in the corresponding modal logic, we have*

$$\{X_1 \leftrightarrow \Box X_1\}_{i \in I} \vdash P \tag{5}$$

*for some set of non-modal propositions  $X_i$ .*

*Proof.*  $P$  is causally entailed according to  $\Xi$  iff  $P$  is true in each causally entailed model, i.e. if

$$P \in \bigcap \{\mathcal{I} \mid \mathcal{I} \text{ a causally explained model}\};$$

but, by the previous lemma,

$$\bigcap \{\mathcal{I} \mid \mathcal{I} \text{ a causally explained model}\} = \overline{\{P \leftrightarrow \Box P \mid P \text{ non-modal}\}}$$

The result follows.  $\square$

This allows us a more intuitive description of the concept of a causally explained model. For a proposition  $P$ ,  $\Box P$  is true in a model if we have an explanation of  $P$ , true in that model: so, for a model to be causally explained, two things must hold. Firstly, every proposition true in the model must have an explanation, true in that model; secondly, every proposition with an explanation, true in the model, must itself be true in the model.

### 3 Explanation and Argument

Parsons and Jennings ([8]; see also [9]) have described a consequence relation,  $\vdash_{\text{ACR}}$ , which is intended to capture the practice of argumentation. The items which this system manipulates are ordered pairs  $(p, A)$ , where  $p$  is a proposition and  $A$  a set of propositions: we will call such a pair an *argument*. Intuitively,  $p$  is the conclusion of an argument, and  $A$  is the set of its grounds. We will also call  $p$  the *head* of the argument, and  $A$  its *body*.

The system is given in Table 2: here  $\Xi$  is a set of *basic* arguments. Parsons and Jennings write their system in natural deduction style, with introduction and elimination rules, using sequents of the form  $\Xi \vdash_{\text{ACR}} (p, A)$ : such a sequent says that  $A$  is an argument for  $p$ , given basic arguments  $\Xi$ .

(Note that we have interchanged the labels on the rules  $\neg$ I and RAA: Parsons and Jennings' original labelling is clearly a typo of some sort.)

We can now translate Parsons and Jennings' system into ours.

**Definition 1.** Let  $\Xi \vdash_{\text{ACR}} (p, A)$  be a sequent in Parsons and Jennings' system. Its *modal translation* is the sequent

$$A \vdash \Box_{\Xi} P,$$

where  $\Box_{\Xi}$  is the modal operator defined by causal laws

$$\left\{ \left( \bigwedge_{a \in A} a \right) \triangleright p \mid (p, A) \in \Xi \right\}.$$

Since the Parsons and Jennings system is written in natural deduction style, some of the rules (for example  $\rightarrow$  I) manipulate the set of basic arguments: consequently, the modal operator in the modal translation will vary. We will, then, need the following lemma:

**Lemma 4.** *If  $\Box_\Xi$  is the modality associated to a set  $\Xi$  of explanations, and if  $\Box_{\Xi \cup \{(\psi, \phi)\}}$  is that associated to  $\Xi \cup \{(\psi, \phi)\}$ , then we have*

$$\Box_{\Xi \cup \{(\psi, \phi)\}} p \cong (\phi \wedge \Box_\Xi(\psi \rightarrow p)) \vee \Box p$$

for any  $p$ .

*Proof.* This follows easily using Proposition 1: we show that operator on the right hand side (i.e.  $\phi \wedge \Box_\Xi(\psi \rightarrow \cdot) \vee \Box \cdot$ ) is a **K** modality, and satisfies the same minimality property as  $\Box_\Xi$ . The two are, consequently, equivalent.  $\square$

We have the following

**Proposition 4.** *The modal interpretation is sound: that is, each of Parsons and Jennings' axioms is translated into a tautology.*

*Proof.* **Ax** is

$$\frac{}{\Xi \vdash_{\text{ACR}} (p, A)} (p, A) \in \Xi$$

and this follows from our definition of the modal translation.

$\wedge$ -**I**,  $\neg$ -**E**,  $\rightarrow$ -**E**  $\wedge$  - **I**, for example, is

$$\frac{\Xi \vdash_{\text{ACR}} (p, A) \quad \Xi \vdash_{\text{ACR}} (q, B)}{\Xi \vdash_{\text{ACR}} (p \wedge q, A \cup B)}$$

and this follows from the **K** tautology  $\Box p \wedge \Box q \vdash \Box p \wedge q$ .  $\neg$  - **E** and  $\rightarrow$  - **E** are similar.

$\wedge$ -**E1**,  $\wedge$ -**E2**,  $\vee$ -**I1**,  $\vee$ -**I2**, **EFQ**  $\wedge$  - **E1**, for example, is

$$\frac{\Xi \vdash_{\text{ACR}} (p \wedge q, A)}{\Xi \vdash_{\text{ACR}} (p, A)}$$

and this follows from the **K** tautology  $\vdash \Box(a \wedge b) \rightarrow \Box a$ .  $\wedge$  - **E2**,  $\vee$  - **I1**,  $\vee$ -**I2**, and **EFQ** are similar.

$\top$ -**I** This is just

$$\vdash \Box \top,$$

a **K** tautology.

$\vee$ -**E** This is

$$\frac{\Xi \vdash_{\text{ACR}} (p \vee q, A) \quad \Xi, (p, A) \vdash_{\text{ACR}} (r, B) \quad \Xi, (q, A) \vdash_{\text{ACR}} (r, C)}{\Xi \vdash_{\text{ACR}} (r, B \cup C)}$$

and this corresponds, in our system, to

$$A \rightarrow \Box_\Xi(p \vee q), B \rightarrow \Box_{\Xi \cup \{(p, A)\}} r, C \rightarrow \Box_{\Xi \cup \{(q, A)\}} r \\ \vdash B \wedge C \rightarrow \Box_\Xi r$$



We can use the lemma to express  $\Box_{\Xi \cup \{(p,A)\}}$  and  $\Box_{\Xi \cup \{(q,A)\}}$  in terms of  $\Box_{\Xi}$ ; some routine but tedious computation then reduces this case to

$$\Box_{\Xi}(p \vee q), \Box_{\Xi}(p \rightarrow r), \Box_{\Xi}(q \rightarrow r) \vdash \Box_{\Xi}r \quad (6)$$

which is a **K** tautology.

**RAA**,  $\neg$ -**I**  $\neg$ -**I** is

$$\frac{\Xi, (p, \emptyset) \vdash_{\text{ACR}} (\perp, A)}{\Xi \vdash_{\text{ACR}} (\neg p, A)}$$

which corresponds to

$$A \rightarrow \Box_{\Xi \cup \{(p, \emptyset)\}} \perp \vdash A \rightarrow \Box_{\Xi} \neg p;$$

using the lemma on  $\Box_{\Xi \cup \{(p, \emptyset)\}}$ , and some computation, reduces this to

$$\Box \perp \vee \Box (p \rightarrow \perp) \vdash \Box \neg p$$

which is a **K</**

**Definition 2.** Let *classical*  $\vee E$  be the following rule:

$$\frac{A \vdash q \vee r \quad \Xi \vdash_{\text{ACR}} (r, B \cup \{q\}) \quad \Xi \vdash_{\text{ACR}} (r, C \cup \{r\})}{\Xi \vdash_{\text{ACR}} (r, A \cup P \cup C)} \vee E$$

where  $A \vdash q \vee r$  is an entailment in classical natural deduction.

We clearly have

**Proposition 6.** *The modal translation is sound for classical or-elimination.*

And we can also prove completeness:

**Theorem 4.** *The modal translation is complete: that is, given a proof of*

$$A \vdash \Box_{\Xi} p, \tag{7}$$

*there is a proof, in the Parsons and Jennings system together with classical or-elimination, of*

$$\Xi \vdash_{\text{ACR}} (p, A) \tag{8}$$

which is valid: that is, from  $\Xi \vdash_{\text{ACR}} (p, A)$  and  $\Xi', (p, A) \vdash_{\text{ACR}} (q, B)$ , we can establish  $\Xi', \Xi \vdash_{\text{ACR}} (q, B)$ . But this is not enough: we *also* want to regard the grounds of arguments as premises.

The situation is clearly two-dimensional in something like Pratt's sense – he defines the *dimension* of a logic to be “the smallest number of variables and constants of the logic sufficient to determine the remaining variables and constants” [12]: the modal operator can be varied quite independently of the classical connectives, merely by altering the set of causal rules. Consequently, a formalism such as Masini's [5] [6] may well be more appropriate.

**Remark 3.** This translation between sequent calculus and the Parsons and Jennings natural deduction is, in addition, not very sensitive to the structure of proofs on either side: natural deduction proofs tend to transform the conclusion of the argument quite extensively before coming down to basic arguments. Sequent calculus proofs, by contrast, leave the conclusion unchanged until an application of  $\Box R$ , after which the proof is a matter of standard classical logic. In addition, the Parsons and Jennings system only represents a fragment of the full sequent calculus (namely, the entailments in which  $\Box$  only occurs on the right). A natural deduction formulation of the entire sequent calculus would be interesting, but would have to extend the Parsons and Jennings system quite considerably.

## References

- [1] Nuel D. Belnap and Thomas B. Steel. *The Logic of Questions and Answers*. Yale UP, New Haven, CT, 1976.
- [2] Matt Fairtlough and Stanley S. Wainer. Hierarchies of provably recursive functions. In Samuel R. Buss, editor, *Handbook of Proof Theory*, volume 137 of *Studies in Logic and the Foundations of Mathematics*, pages

- [8] Simon Parsons and N. R. Jennings. Negotiation through argumentation: A preliminary report. In *Proceedings of the International Conference on Multi Agent Systems (ICMAS'96)*, Kyoto, Japan, 1996.
- [9] Simon Parsons, Carles Sierra, and Nick Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.
- [10] A. Pliuškevičienė. Extended disjunction and existence properties for some predicate modal logics. *Logic Journal of the IGPL*, 6(5):775–787, 1998.
- [11] Wolfram Pohlers. *Proof Theory: An Introduction*. Number 1407 in Lecture Notes in Mathematics. Springer, Berlin etc., 1989.
- [12] Vaughan Pratt. A roadmap of some two-dimensional logics. In J. van Eijck and A. Visser, editors, *Proc. Workshop on Logic and Information Flow*, pages 149–162. MIT Press, 1993.
- [13] Peter Schroeder-Heister. Cut-elimination in logics with definitional reflection. In D. Pearce and H. Wansing, editors, *Nonclassical Logics and Information Processing*, volume 619 of *Lecture Notes in Artificial Intelligence*, pages 146–171, Berlin, etc., 1992. Springer-Verlag. International Workshop, Berlin 1990.
- [14] A.S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Number 43 in Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 1996.
- [15] Hudson Turner. A theory of universal causation. *Artificial Intelligence*, 113:87–123, 1999.
- [16] G. Graham White. A modal formulation of McCain and Turner’s theory of causal reasoning. In Sergio Flesca, Sergio Greco, Nicola Leone, and Giovambattista Ianni, editors, *Logics in Artificial Intelligence: 8th European Conference, JELIA 2002*, number 2424 in Lecture Notes in Artificial Intelligence, pages 211–222. Springer, 2002.
- [17] Ludwig Wittgenstein. *Philosophical Investigations*. Blackwell, 1967.

Table 2: The Parsons and Jennings System

---

$\frac{(p, A) \in \Xi}{\Xi \vdash_{\text{ACR}} (p, A)} \text{Axiom}$	$\frac{}{\Xi \vdash_{\text{ACR}} (\top, \emptyset)} \top\text{I}$
$\frac{\Xi \vdash_{\text{ACR}} (p, A) \quad \Xi \vdash_{\text{ACR}} (q, B)}{\Xi \vdash_{\text{ACR}} (p \wedge q, A \cup B)} \wedge\text{I}$	$\frac{\Xi \vdash_{\text{ACR}} (p \wedge q, A)}{\Xi \vdash_{\text{ACR}} (p, A)} \wedge\text{E1}$
	$\frac{\Xi \vdash_{\text{ACR}} (p \wedge q, A)}{\Xi \vdash_{\text{ACR}} (q, A)} \wedge\text{E2}$
$\frac{\Xi \vdash_{\text{ACR}} (p, A)}{\Xi \vdash_{\text{ACR}} (p \vee q, A)} \vee\text{I1}$	$\frac{\Xi \vdash_{\text{ACR}} (p \vee q, A) \quad \Xi, (p, A) \vdash_{\text{ACR}} (r, B) \quad \Xi, (q, A) \vdash_{\text{ACR}} (r, C)}{\Xi \vdash_{\text{ACR}} (r, B \cup C)} \vee\text{E}$
$\frac{\Xi \vdash_{\text{ACR}} (q, A)}{\Xi \vdash_{\text{ACR}} (p \vee q, A)} \vee\text{I2}$	
$\frac{\Xi, (p, \emptyset) \vdash_{\text{ACR}} (\perp, A)}{\Xi \vdash_{\text{ACR}} (\neg p, A)} \neg\text{I}$	$\frac{\Xi \vdash_{\text{ACR}} (p, A) \quad$